ENGINEERED OLIGOSACCHARYLTRANSFERASES WITH UNIQUE N-GLYCOSYLATION SITE PREFERENCES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Feng-Yang Chen

August 2019

ABSTRACT

The key enzyme in the *Campylobacter jejuni* glycosylation pathway, PglB, has been one of the well-studied asparagine-linked oligosaccharyltransferase (OST) in bacterial glycoengineering area. While *C. jejuni* PglB performs well in transferring different glycans, the acceptor sites it can recognize are limited. We provided a thorough investigation into *Desulfovibrio desulfuricans* PglB (DdPglB), an OST sharing homology with *C. jejuni* PglB but having more relaxed sequon specificity. The sequon logo of DdPglB against XXNXT library were built through a high-throughput assay. Two engineered OSTs combining distinct transmembrane domain and periplasmic domain of PglBs from *C. jejuni* and *D. desulfuricans* was also investigated. And the study interestingly revealed they can lose all their glycosylation activities after the combination. According to the sequon logo we built and further verification, a notable sequon QYNST, which is a native acceptor sequon of Fc domain of human immunoglobulin G, was able to be glycosylated by DdPglB. Overall, we presented a series of sequons that can be glycosylated by DdPglB and further expanded the glycoengineering toolbox with more potential acceptor sites that were not glycosylated before.

BIOGRAPHICAL SKETCH

Feng-Yang Chen received his B.S. in Chemical Engineering at National Taiwan University in 2016 and completed an industry-university cooperative research project in a tissue engineering lab in his senior year. After graduation, he entered Cornell University to pursue M.S. degree in Chemical and Biomolecular Engineering and joined DeLisa Research Group as a researcher. During his M.S. degree, he involved in several projects, including *Glyco-mutagenesis of proteins in Escherichia coli* and *Engineered novel oligosaccharyltransferases with relaxed specificity*. He will attend Rice University to pursue a Ph.D. degree in Chemical and Biomolecular Engineering after graduating from Cornell.

This work is dedicated to my family, especially my Mom for her support throughout my education

ACKNOWLEDGMENTS

I would like to thank Professor Matthew DeLisa for his support and advice. I would also like to thank all of the members in DeLisa Research Group, especially Dr. Mingji Li, who taught me almost all the experimental skills in the lab and mentored me for two years during my M.S. degree. Without them, I would not complete this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Protein glycosylation, the covalent modification that attaches oligosaccharides to the specific protein side chain, is one of the most prevalent post-translational modifications found in eukaryotes. In eukaryotes, over 50% of proteins are glycosylated[1]. In addition, diverse glycans and different glycosylation sites could increase glycoproteins' structure complexity and functional variety. Because of these, glycoproteins play important roles in various biological processes and cellular regulation[2].

Importance of glycosylation

Researches have revealed that glycosylation can affect protein properties on different perspectives after modification. For instance, some of the plasma membrane proteins and secretory proteins require the help of glycans to fold properly[3]. Glycosylation also affects the local structure and enhances the overall stability of proteins[4]. Moreover, glycans not only promote protein folding and stability, but also serve as a recognition tag to influence molecular interaction in cellular processes[5]. In addition, studies show that glycosylation is important in influencing protein homeostasis and trafficking[6].

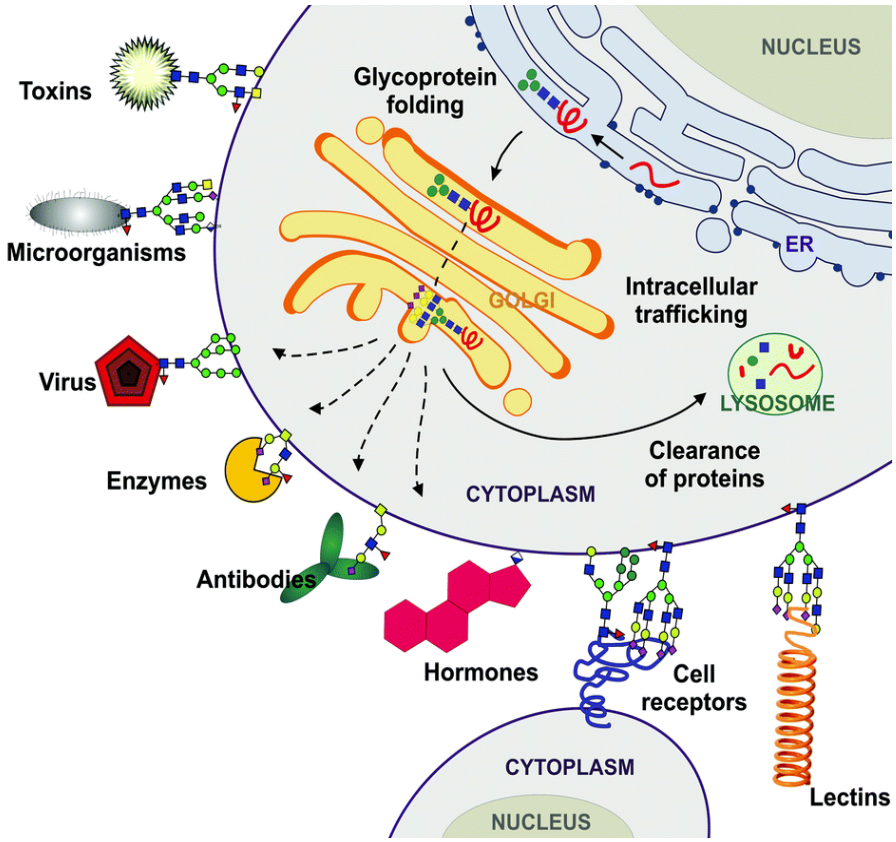

Figure 1. Multiple cellular regulation mechanisms with glycosylation involved[7].

Besides affecting protein properties, glycosylation is also critical in biological regulation. Specific glycans can participate in processes such as immune response to pathogens, adhesion, and cell-cell recognition[8,9]. Furthermore, intentionally modifying protein-associated glycans can be used to enhance the pharmacokinetic and therapeutic properties of proteins, such as activity, half-life, and stability in vivo[10,11]. Schematic in Figure 1 briefly summarizes that various cellular regulations all require the participation of glycosylation process[7].

Types of glycosylation

Glycosylation can happen at several kinds of amino acid side-chains, and glycopeptide bonds can be arranged in distinct groups. Two main types of glycosylation are asparagine-linked (N-linked) glycosylation and O-linked glycosylation, where N-linked glycans covalently link to asparagine or arginine amide nitrogen, and O-linked glycans link to the hydroxyl oxygen of serine, threonine or other amino acid side-chains[12]. Other minor types of glycosylation linkage include C-mannosyl bonds, phosphoglycosyl bonds, and glypiated linkage[13].

Among these types of glycosylation, N-linked is the most universal type. In this study, we focused on N-linked glycosylation to make a deeper investigation.

N-linked glycosylation in eukaryotes and prokaryotes

Although glycosylation was first found in eukaryotes, it was discovered prokaryotes also had native glycosylation machinery in later years. *Campylobacter jejuni*, a pathogenic gram-negative proteobacterium, was the first bacterium found with N-linked glycosylation pathway[14]. The pathway is encoded on the gene locus *pgl*, and locus pglB, which is a part of pgl, has been characterized to be translatable into a protein, oligosaccharyltransferases (OST). Other parts of *pgl*, including glycosyltransferase, flippase and bacillosamine, are responsible for the formation of glycan structure and flipping the glycan outward to the outer membrane[15]. Studies also revealed that the glycan structure of *C. jejuni* is the heptasaccharide $GalNAc_5GlcBac$ by mass spectrum (MS) and nuclear magnetic resonance (NMR)[16].

Glycosylation systems in eukaryotes and prokaryotes share some similar features and mechanics[17]. For example, they both involve the synthesis of a lipid-linked oligosaccharide (LLO) donor, assembling with the glycan structure, transferring across membrane by flippase, and finally transferring to a target protein sequon catalyzed by OST. The whole process can be seen in the schematic in Figure 2.

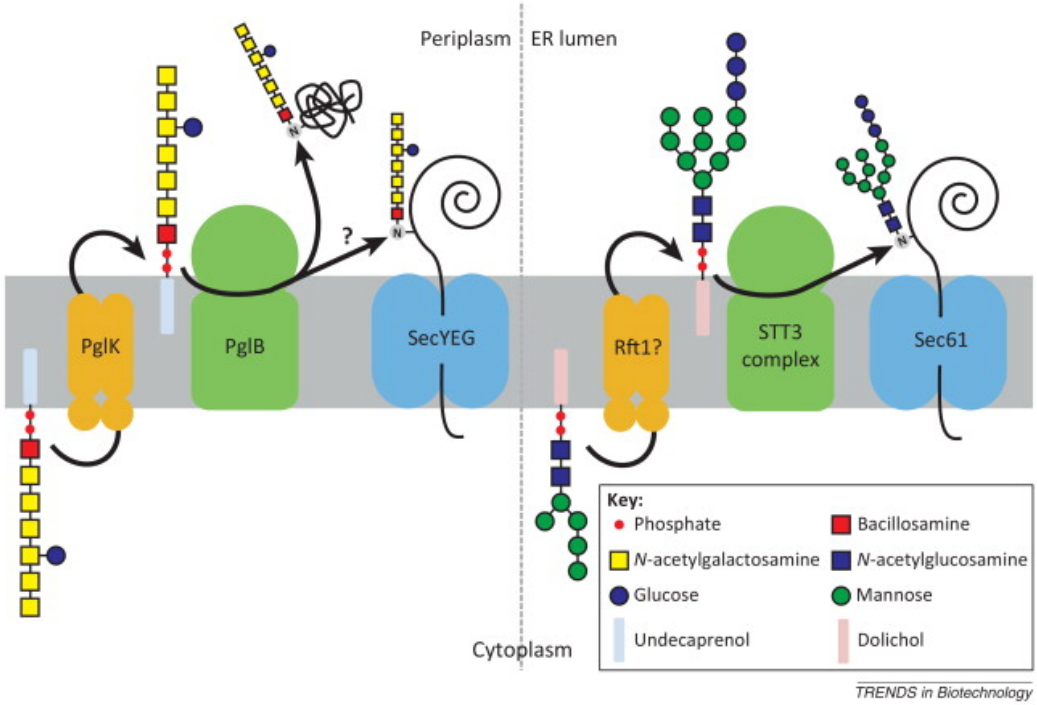

Figure 2. Comparison of N-*linked glycosylation in prokaryotes (left) and eukaryotes (right)*[18].

On the other hand, there are also some notable differences between them. For instance, the oligosaccharides components, the lipid carrier, and the location where LLO assembly are different. The location of protein glycosylation pathway is also distinct. In eukaryotes, the modification occurs in the endoplasmic reticulum (ER) and Golgi, while in prokaryotes, taking C. jejuni as an example, glycan is synthesized on

the cytoplasmic side of the inner membrane and then flipped across the membrane and transferred to the substrate protein by membrane enzyme OST.

OST is another significant difference between these two systems. In *C. jejuni* and most of the bacteria with glycosylation pathway, OST is a monomeric membrane protein named PglB. In contrast, in higher eukaryotes like humans, OST is constituted by several membrane proteins as a complex, with one of the proteins, STT3, acts as the key catalytic subunit[19]. In addition, the acceptor site sequences that OST can recognize are also distinct. More details about the sequences will be given in the next section. Interestingly, prokaryotic PglBs do share significant sequence homology with eukaryotic catalytic unit STT3[20]. Thus we believe, with a deeper study into *C. jejuni* PglB, we can gain more information on eukaryotic glycosylation system.

N-linked glycosylation Consensus sequon

To successfully attach a glycan onto a target protein, it requires specific amino acid sequences for OSTs to recognize. For eukaryotes and archaea, glycosylation occurs to the amide nitrogen located in the sequence N-X-S/T, where X can be any amino acid but proline[21].

On the other hand, in bacteria such as *C. jejuni*, some initial studies found that N-X-S/T consensus sequence is required, but not sufficient for glycosylation[22]. While the later study extended the consensus sequon by a negatively charged amino acid in the -2 position in D/E-X_{-1}-N-X_{+1}-S/T, where X_{-1} and X_{+1} can be any amino acids except proline[23]. This means *C. jejuni* PglB (CjPglB) requires a more stringent motif compared to eukaryotic OSTs. Moreover, the sequon DQNAT was defined to be the

optimal sequon by an experimental way with a series of acceptor peptide substrates and a non-*C. jejuni* model protein, Im7[24]. From these studies, we noticed that OSTs does play the key role in determining which protein can be glycosylated.

Mechanism of oligosaccharyltransferase

Studying the mechanism of OSTs could be the key to understand the whole picture of glycosylation. A previous study provided the X-ray structure of *C. lari* PglB (ClPglB) in complex with an acceptor peptide[25]. From this study and the structure shown in the top picture of Figure 3, we can conclude some critical points. First, this structure provided an insight into glycosylation sequon recognition. For example, the salt bridge between amino acid R331 of the ClPglB and aspartic acid at -2 position of acceptor peptide can strengthen the interaction between PglB and target protein. This might be the reason why CjPglB and ClPglB require D/E at -2 position to glycosylate protein. Another example is the van der Waals contact between the OST and threonine at +2 position of the peptide. This force acts as a stabilizing interaction and also explains why previous studies discovered glycosylation efficiency of threonine at +2 position is much more efficient than serine at +2 position.

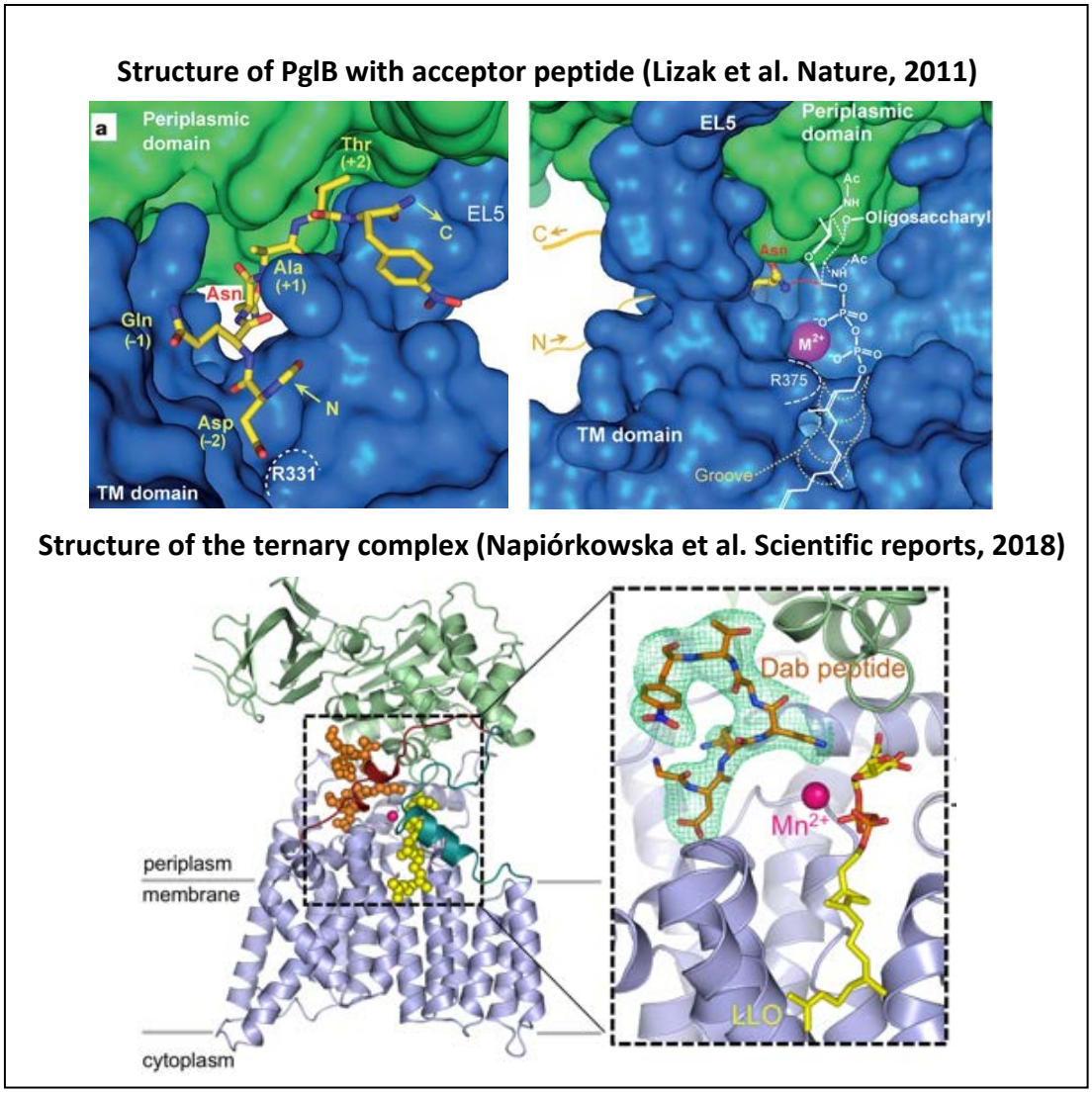

Figure 3. X-ray structure of C. lari PglB. PglB in complex with acceptor peptide DQNATF (top)[25] and ternary complex structure of PglB, acceptor peptide and glycosidic oxygen of LLO (bottom)[26].

Besides, this study also proposed a possible glycosylation mechanism by the structure they observed. Further, several points have been concluded and stated below. First, the function of divalent cation M^{2+} was elucidated. This cation not only orients the acidic residues that interact with the acceptor asparagine, but also stabilizes the leaving LLO and glycan after successfully glycosylation. This also explained why a

cation has been usually observed in this mechanism, and might be necessary for glycosylation process. Next, the periplasmic external loops in the OST structure, EL5, plays an important role in this reaction. EL5 was expected to be a flexible loop in the absence of the acceptor peptide. In the sequon binding stage, part of the EL5 is self-organized and engages in the complex structure to stabilize the periplasmic domain of the OST and the acceptor sequon. After the glycan is attached, the glycopeptide is released by the disengagement of EL5. The later research also supported this proposed mechanism with ternary complex structure[26].

These pieces of evidence have all shown that specific amino acid residues on OSTs are crucial elements that promote glycosylation. Moreover, the transmembrane domain and periplasmic domain of the PglB do involve in the mechanism in different respects, but both play critical roles in the glycosylation process.

Engineering oligosaccharyltransferase

To increase the utilization of glycosylation, as well as broaden the knowledge of glycosylation mechanism, engineering OSTs can be a way with great potential, especially with the help of information of OSTs structure. Engineering OSTs might have the possibility of increasing the glycosylation efficiency of OSTs or making their acceptor-site specificity much broader; therefore, a number of groups have put effort into it. A study from DeLisa research group has successfully engineered CjPglB, and the mutants did have a more relaxed specificity[27]. The idea of engineering came from previous studies mentioned above, that R331 residue of ClPglB was discovered to form a salt bridge between OST and -2 position of acceptor peptide. Thus, this study

modified the residues of CjPglB which is conserved between *C. jejuni* and *C. lari*, and the engineered OSTs do have a wilder specificity against -2 position of the acceptor sequon. Furthermore, these engineered OSTs do have the ability to glycosylate wild-type RNase A, which could not be glycosylated in the past by native CjPglB.

Another study also focused on certain residues of CjPglB and randomly engineered these residues to examine if it is possible to pick up OST mutants with higher glycosylation efficiency[28]. After screening and selection by ELISA assay against mutagenesis library, several mutants have been picked up. The results showed that compared to wild-type CjPglB, some of the mutants did have better ability to glycosylate specific proteins, that is, the performance on glycosylation efficiency of the OST mutants increased significantly.

These examples demonstrated that by engineering and modifying certain amino acid of OSTs, there is potential to increase their glycosylation efficiency. Moreover, it is also possible to create an engineered OST with relaxed sequon specificity, which means we will be able to glycosylate more proteins that we were not able to glycosylate in the past, and these new glycoproteins could further raise the potential for the pharmaceutical industry.

Protein glycosylation in bacteria

Instead of *C. jejuni*, there are more bacteria found to have similar glycosylation pathway. Some of these bacteria's Pgl sequences do share some homology with *C. jejuni*, while the glycosylation efficiency and sequon specificity could be quite distinct. Research from DeLisa group has provided a thorough investigation into specificity for

the -2 sequon residue of several bacteria's PglBs, which have native glycosylation pathway[29]. The results are shown in Figure 4. From the results, we can notice that as what we knew before, *C. jejuni* and *C. lari* PglB do have a narrow specificity (required D and E at -2 position of the sequon) but a pretty high glycosylation efficiency, which is 100% here.

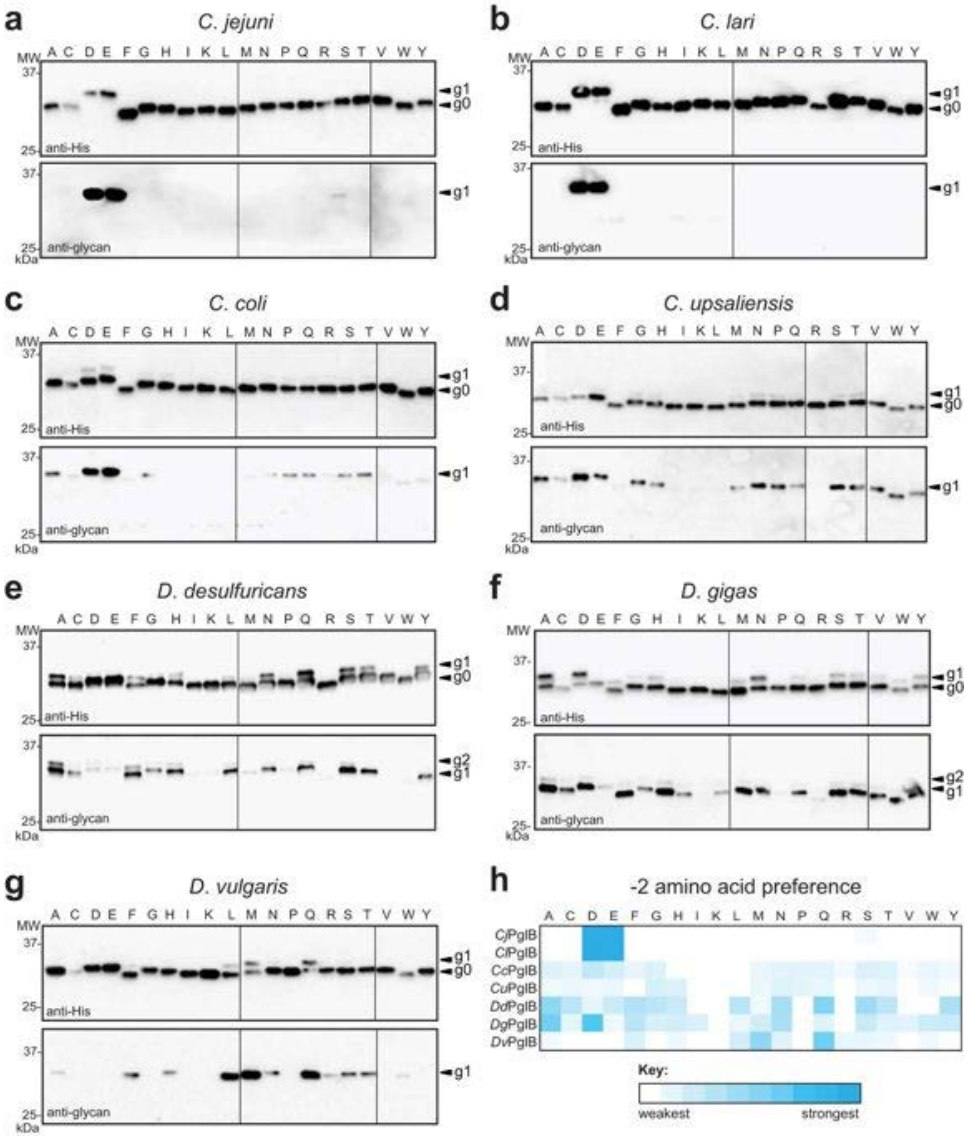

Figure 4. Glycosylation of bacteria's PglB against scFv13-R4^{XQNAT} variants (where X = one of the 20 amino acids)[29].

In addition, *Desulfovibrio desulfuricans* is an interesting target here. *D. desulfuricans* PglB (DdPglB) has a quite relaxed specificity compared to all the other targets. Although its glycosylation efficiency is not as high as CjPglB and ClPglB, its performance is still very good compared to other targets, which also have relaxed specificities. We believe DdPglB has a great potential to glycosylate more important sequons that have not been glycosylated by other OSTs before. If we can properly modify DdPglB, it might become an engineered OST with relaxed specificity and high glycosylation efficiency. Therefore, we have chosen DdPglB as our main target OST to study in this research.

Studying glycosylation in E. coli

One of the major steps for glycoengineering is functionally transferring the *C. jejuni* glycosylation system into *E. coli*[30]. The bacterium *E. coli* is a good platform for some reasons. First, *E. coli* is a well-studied bacterium with the ease of production and manipulation. Besides, *E. coli* does not have glycosylation pathways; thus, we do not need to worry about if transferring the glycosylation system into *E. coli* will affect either the glycosylation system or *E. coli*. Also, *E. coli* is suitable for protein production and is a well-known bioengineering system, hence it has the potential to produce glycoengineering products.

In this research, we used E. coli as our system to investigate the unknown sequon preferences of OSTs and to see if we could use engineered OSTs to glycosylate the protein of interest.

Research objectives

As mentioned above, DdPglB has quite relaxed specificity; thus, we believed it also has the potential to glycosylate more sequons that have never been glycosylated before. Therefore, the first aim of this study is to investigate the sequon logo of DdPglB against XXNXT sequon library, where X is a randomly modified amino acid. Here we used Im7[24], a model protein which has been used to study *C. jejuni* glycosylation before, as our target protein. We also introduced a high-throughput screening assay developed by DeLisa group, named glycoSNAP (*glyco*sylation of *s*ecreted *N*-linked *a*cceptor *p*roteins)[31], to complete the sequon logo of DdPglB.

The second objective of this research is to engineer DdPglB to see if it is possible to create mutants that keep relaxed specificity but have higher glycosylation efficiency compared to wild-type DdPglB. From the researches mentioned above, we noticed that the transmembrane domain and periplasmic domain of PglB might act functionally distinctly on stabilizing specific substrates or glycosylating certain target proteins. Accordingly, this research created chimeric OSTs that can combine different domains of CjPglB, which have high glycosylation efficiency, with DdPglB, which have relaxed specificity, and see the sequon preference and performance of the chimeric OSTs against sequons with different -2 position amino acid.

Many important proteins have native conserved N-linked glycosylation site, but some of them are still left to be well-glycosylated. Therefore, the third aim of this project is to see if DdPglB can glycosylate certain wild-type protein's native glycosylation site in our bacteria system. The investigation of DdPglB against the XXNXT sequon can be used here to select the most promising target, and can further

test the glycosylation performance of *D. desulfuricans* OST against this target of interest. The result of this study provided a potential glycoengineering target and a glycoprotein that can be engineered by DdPglB.

CHAPTER 2

MATERIALS AND METHODS

Overview

To reach all the goals of this study, *E. coli* was used to perform all of the glycosylation. Three main plasmids which encoded the glycosylation pathway were transformed into E. coli, and the three plasmids stands for glycan production pathway, OST, and acceptor protein respectively. After the transformation, plasmids in *E. coli* can be induced by inducers to express the protein and execute glycosylation. Western blot analysis was used to detect the aglycosylated and glycosylated proteins and their efficiency. GlycoSNAP assay was used to screen the acceptor protein library for the first aim of this project. Schematic in Figure 5 briefly shows the process of this study.

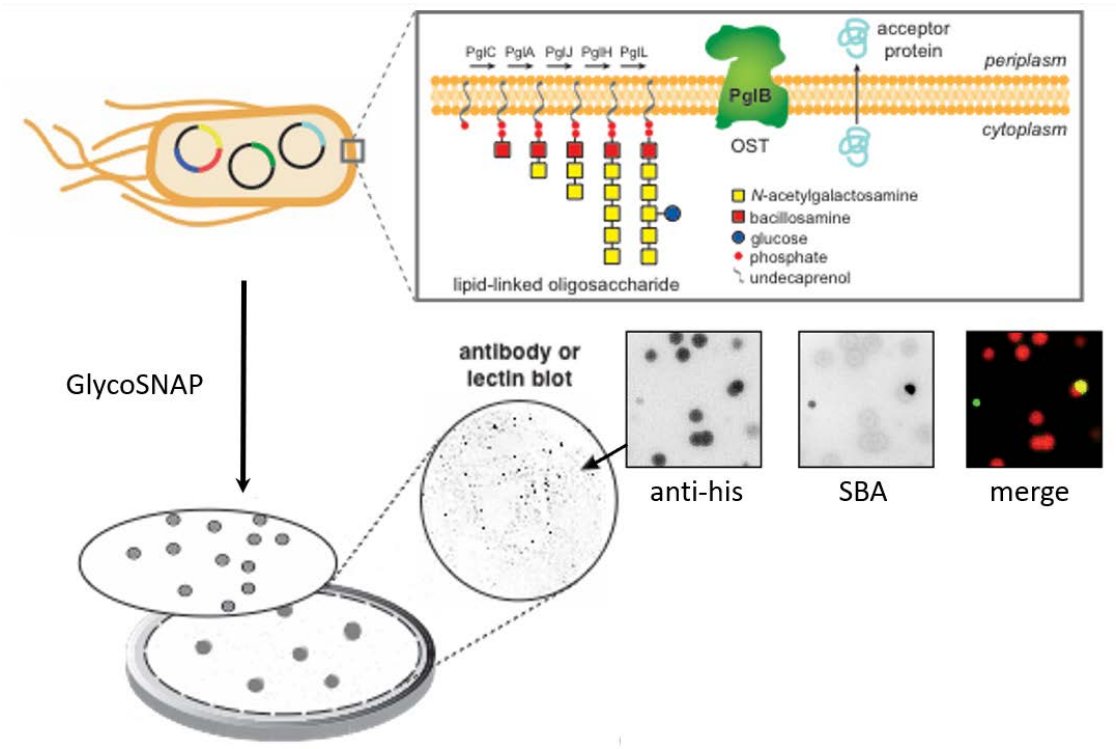

Figure 5. Overview of the glycosylation pathway in E. coli followed by glycoSNAP.

Bacterial strains and growth conditions

The following *E. coli* strains were used in this study: DH5α and CLM24. Details are listed in Table 1 below. DH5α was used for all of the plasmid cloning, site-directed mutagenesis, and library construction. CLM24 was used for all *in vivo* glycosylation studies. Cultures were grown in Luria-Bertani (LB) broth at 37 °C containing antibiotics as needed. Concentrations of the antibiotics are listed as following: 100 μg/ml ampicillin (Amp), 20 μg/ml chloramphenicol (Cm), 100 μg/ml trimethoprim (Tp), and 50 μg/ml spectinomycin (Spec). For protein expression, cultures were typically induced at OD_{600} ~0.8 by 0.04% (w/v) L-arabinose (L-ara) for pBAD-based expression vectors and 0.1 mM isopropyl β-D-thiogalactoside (IPTG) for pTrc99-based expression vectors. Induction was carried out at 30 °C for 20 h. Condition of protein expression was determined by a series of optimization shown in Figure S1, S2, and S3 in Appendix.

Table 1. Strains used in this study

Strains	Relevant genotype	Reference or source
DH5α	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ^-	Laboratory stock
CLM24	*E. coli* W3110 *ΔwaaL*	Feldman et al.[32]

Plasmid construction

All of the plasmids used in this study are listed in Table 2. Details of cloning and plasmid construction are described in the following page.

Table 2. plasmids used in this study

Plasmids	Description	Reference
pMW07-pglΔB	*C. jejuni* pgl locus with deletion of CjPglB, Cm^r	Ollis et al.[27]
pMAF10	Wild-type *C. jejuni* PglB, Tp^r	Feldman et al.[32]
pMAF10-$CjPglB^{D54N\ E316Q}$	*C. jejuni* PglB with inactivating mutation, Tp^r	Ollis et al.[27]
pMAF-DdOST	pMAF10- derivative with *C. jejuni* PglB replaced by *D. desulfuricans* PglB, Tp^r	Lab stock
pMAF-ClOST	pMAF10- derivative with *C. jejuni* PglB replaced by *C. lari* PglB, Tp^r	Lab stock
pMAF-CcOST	pMAF10- derivative with *C. jejuni* PglB replaced by *C. coli* PglB, Tp^r	Lab stock
pMAF-$Cj^{TM}Dd^{PD}$OST	pMAF plasmid encoding transmembrane domain of CjPglB attached with periplasmic domain of DdPglB, Tp^r	This study
pMAF-$Dd^{TM}Cj^{PD}$OST	pMAF plasmid encoding transmembrane domain of DdPglB attached with periplasmic domain of CjPglB, Tp^r	This study
pTrc99S-YebF-$Im7^{DQNAT}$	pTrc plasmid encoding signal peptide and model protein YebF-Im7 with engineered glycosylation tag DQNAT at C-terminus, Sp^r	Lab stock
pTrc99S-YebF-$Im7^{QYNST}$	Site-directed mutagenesis of pTrc99S-YebF-$Im7^{DQNAT}$ with QYNST sequon, Sp^r	This study
pTrc99S-YebF-$Im7^{XQNAT}$	Series of 20 separated plasmid constructed by site-directed mutagenesis of pTrc99S-YebF-$Im7^{DQNAT}$ with XQNAT sequon, where X are 20 distinct amino acid, Sp^r	This study
pTrc99S-YebF-$Im7^{XXNXT}$	A mixed plasmid library constructed by site-directed mutagenesis of pTrc99S-YebF-$Im7^{DQNAT}$ with XXNXT sequon, Sp^r	This study

Plasmids used in this study can be separated into three main groups: glycan production pathway, OST, and acceptor protein. For producing glycan, pMW07-pglΔB was used. This plasmid encodes the *C. jejuni pgl* locus with a complete in-frame deletion of *pglB*, thus serves as the whole glycosylation pathway that can produce *C. jejuni* glycan (CJG) without the functionality of OST.

Next, pMAF10 was used to express CjPglB in this study. Plasmid pMAF10-CjPglB$^{D54N\ E316Q}$ was a mutant with inactivation of CjPglB, and thus served as the negative control. pMAF-DdOST was the main target we investigated, while pMAF-ClOST and pMAF-CcOST were also used as preliminary tests and comparison. Plasmids pMAF-Cj^{TM}Dd^{PD}OST and pMAF-Dd^{TM}Cj^{PD}OST were separately constructed by cloning transmembrane and periplasmic domains of CjPglB and DdPglB into pMAF backbone. PCR fragments of backbone, CjPglB, DdPglB were produced and then followed by Gibson assembly to complete the plasmid construction. Crossover sequence boundary of transmembrane and periplasmic domains were determined by homology comparison of the crystal structure of *C. lari* PglB[25] and a proteins' transmembrane helices prediction model[33].

Plasmid pTrc99S-YebF-Im7^{DQNAT} was used as the optimal acceptor protein for CjPglB. This plasmid encodes the native *yebF* gene from *E. coli*, which can be translated into a signal peptide that makes *E. coli* transfer the target protein attached at its C terminus out of the outer membrane. Im7 is a model protein which has been used to study glycosylation of DQNAT sequon before. The engineered glycosylation sequence right after Im7, DQNAT, is the optimal acceptor sequon for CjPglB[24]. Plasmids pTrc99S-YebF-Im7^{QYNST} and pTrc99S-YebF-Im7^{XQNAT} were generated by

replacing DQNAT sequon with QYNST and XQNAT (X stands for 20 distinct amino acid) sequon by site-directed mutagenesis of the parental pTrc99S-YebF-Im7^{DQNAT} plasmid. Designed primer and PCR were used and followed by Gibson assembly to complete the cloning. Plasmid library pTrc99S-YebF-Im7^{XXNXT} was created by replacing DNA sequence GACCAaacCGCGacc (which can be translated into DQNAT) from pTrc99S-YebF-Im7^{DQNAT} with NNKNNKaacNNKacc (which can be translated into XXNXT). Designed primer set with random NNK sequences were used to perform PCR and followed by T4 ligation and transformation into DH5α. Colonies were then screened onto plates and collected to complete the plasmid library construction.

All plasmids constructed in this study were verified through DNA sequencing at the Cornell Biotechnology Resource Center.

Protein purification

To analyze YebF-Im7 products of *in vivo* glycosylation, supernatant and periplasmic fractions were collected and used for different parts of the experiment. For proteins in the supernatant, cells were pelleted after induced for 20 h and supernatant was collected. For periplasmic fraction, the pellet of cells was collected and resuspended by adding 0.4 M arginine. After reacting in 4 °C for 1 h, the supernatant was collected and stored as proteins in periplasmic fractions. All protein samples were solubilized in Laemmli sample buffer containing 5% β-mercaptoethanol (βME), and then boiled at 95 °C for 10 min and analyzed by western blotting.

Western blot assay

Protein samples were loaded into 10% SDS-PAGE gels (BioRad) and followed by electrophoretic separation at 150 V for 65 min. Proteins were then transferred from gels onto nitrocellulose membrane and incubated for 30 min in blocking solution (5% non-fat milk in TBST). Membranes were washed 3 times with TBST with 10 min incubation between each wash. One of the membranes was then probed with anti-His antibodies (6xHis-polyclonal antibody, Abcam, ab137839, 1:5000 in TBS). Another replicate membrane was probed with hR6 (1:10000 in TBS), a serum from rabbit which can recognize the native *C. jejuni* glycan. The second membrane was then incubated with antirabbit-HRP secondary antibody after washing 3 times with TBST. Membranes were probed with antibodies at room temperature for 1 h with shaking, and then washed with TBST for 3 times with 10 min incubation between each wash. Membranes were then incubated with Western ECL substrate (BioRad). ChemiDocTM XRS+System was used to detect the chemiluminescent signal of the membranes.

GlycoSNAP assay

The target combinatorial library (which is YebF-Im7^{XXNXT} in this study) transformants were plated on 150 mm LB agar plates containing 20 μg/ml Cm, 100 μg/ml Tp as well as 50 μg/ml Spec, and incubated at 37 °C for 16-20 h. Colonies were then replicated onto Whatman 0.45 μm 142 mm cellulose nitrate membrane filters (VWR), which were overlaid on a nitrocellulose transfer membrane (Fisher Scientific). These membranes were placed on induction LB agar plates containing

20 μg/ml Cm, 100 μg/ml Tp, 50 μg/ml Spec, 0.04% (w/v) L-ara, as well as 0.1mM IPTG and were incubated at 30 °C for 20 h. The induction condition was determined by a series of optimization shown in Figure S4 in Appendix. Next day, the filters containing colonies were transferred onto fresh LB agar plates with the same antibiotics and stored at 4 °C for backup. The nitrocellulose membranes were washed in TBS for 10 min and then blocked with 5% (w/v) bovine serum (BSA) in TBST for 30 min. After blocking, the membranes were incubated with lectin soybean agglutinin (SBA, 0.5 μg/mL in TBS) for 1 h, washed by TBS for twice with 10 min incubation between each wash, and then ChemiDocTM XRS+System was used to detect their chemiluminescent signal. Examples of probing results are presented in Figure S4 in Appendix. Colonies with positive signal were then picked from plates and restreaked on new plates with the same antibiotics for subsequent analysis.

Generation of sequon logo

After picking positive hits from GylcoSNAP of DdPglB against YebF-Im7^{XXNXT} library, western blot assay was used to confirm they can be glycosylated. These positive hits were then confirmed by DNA sequencing, and the sequon logo was then generated from our sequencing data by using WebLogo 3[34].

CHAPTER 3

RESULTS

Optimization of GlycoSNAP and growth conditions

Before we performed glycoSNAP assay and further investigated into the sequon logo of DdPglB, we first optimized the whole process. The expression condition from previous study[27] was not suitable for us here, because DdPglB shows a lower glycosylation level compared to CjPglB generally. We used a set of conditions with different amount of inducer or incubation time to find out the optimal condition, and parts of the testing results are shown in Figure 6. From the figure, we could notice that glycoSNAP signals with original condition, 0.2% L-ara plus 0.1mM IPTG, were very weak compared to those with our final condition, 0.04% L-ara plus 0.1mM IPTG. Moreover, we found out the signal strength could be increased with increasing incubation time. Thus, after testing with different sets of conditions, we finally decided to incubate the glycoSNAP at 30 °C for 20 h, and then to incubate at room temperature for 20 h.

Besides the optimization of the glycoSNAP, in order to get bands with high quality on Western blot, we also optimized the growth conditions for DdPglB against YebF-Im7 protein. We tested the conditions with different amount of inducer, incubation time, level of OD_{600}, or protein purification methods. All of the optimization blots are shown in Figure. S1, S2, and S3 in Appendix. Finally, we found out the optimal condition for glycosylation of DdPglB against YebF-Im7 protein, which was inducing the target strains with 0.04% L-ara and 0.1mM IPTG for 20 h when the OD_{600} of *E. coli* in LB is around 0.8. We noticed inducing at higher OD did

not guarantee a better glycosylation performance. In an appropriate range, a higher OD could lead to higher protein expression level; therefore, the signal of glycosylation band (g1 band) could increase, while the aglycosylation band (g0 band) could also increase. Overall, this could cause a slight decrease in glycosylation efficiency.

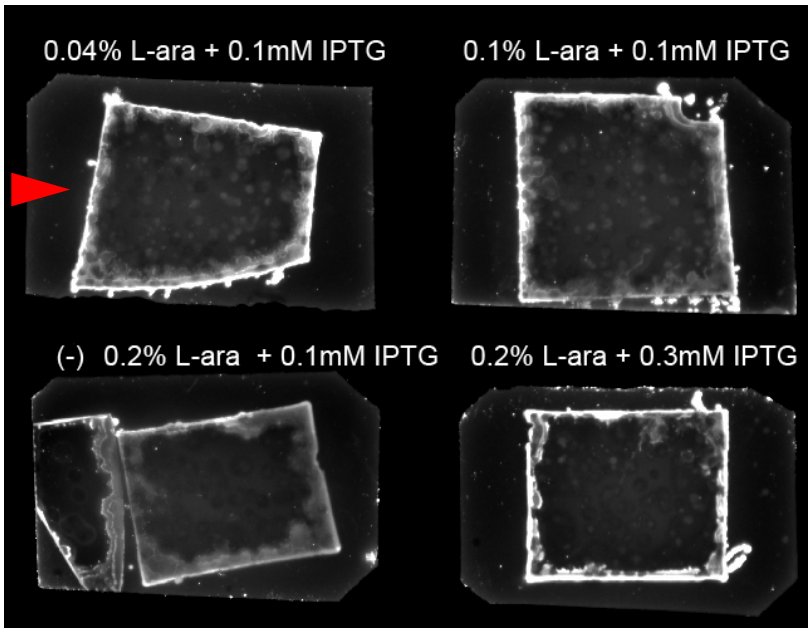

Figure 6. Optimization of glycoSNAP at different conditions. GlycoSNAP of DdPglB against YebF-Im7^{AQNAT} was performed with different amount of inducer, L-ara or IPTG, at 30 °C for 20 h, and then followed by incubating at room temperature for 20 h. All the blots were incubated in the same LB agar plate with appropriate antibiotics and probed with the lectin SBA. The negative control reaction (-) was performed using the inactivated OST mutant CjPglB$^{D54N\ E316Q}$ against YebF-Im7^{DQNAT}. Red arrow indicates the optimal condition we used in this project.

Sequon logos of DdPglB

To fully investigate the acceptor sequon specificity of *D. desulfuricans* PglB, an acceptor protein library YebF-Im7^{XXNXT} was built. Asparagine at the central position was fixed for N-linked glycosylation, and threonine at +2 position was also fixed because it is one of the required residues for eukaryotic OSTs. X at -2, -1 and +1 positions of the sequon represents a randomly generated amino acid. DdPglB was tested against this protein library in *E. coli*, and screened by glycoSNAP assay. 72 potential hits from the glycoSNAP assay were picked up, and further verified by Western blot. 33 hits among them were positively glycosylated. These glycosylated hits were then examined by DNA sequencing to verify the sequon. The Western blot and sequencing results are shown in Figure 7.

The negative control reaction (-) was performed using the inactivated OST mutant CjPglB$^{D54N\ E316Q}$ against YebF-Im7^{DQNAT}, and positive control reaction (+) was performed by DdPglB against YebF-Im7^{AQNAT}, one of sequons that can be glycosylated with the highest efficiency for DdPglB in the previous study[29]. These controls were used to demonstrate the position of the aglycosylated and glycosylated bands in Western blot. The same control reactions were used in all of the following Western blot results in this study.

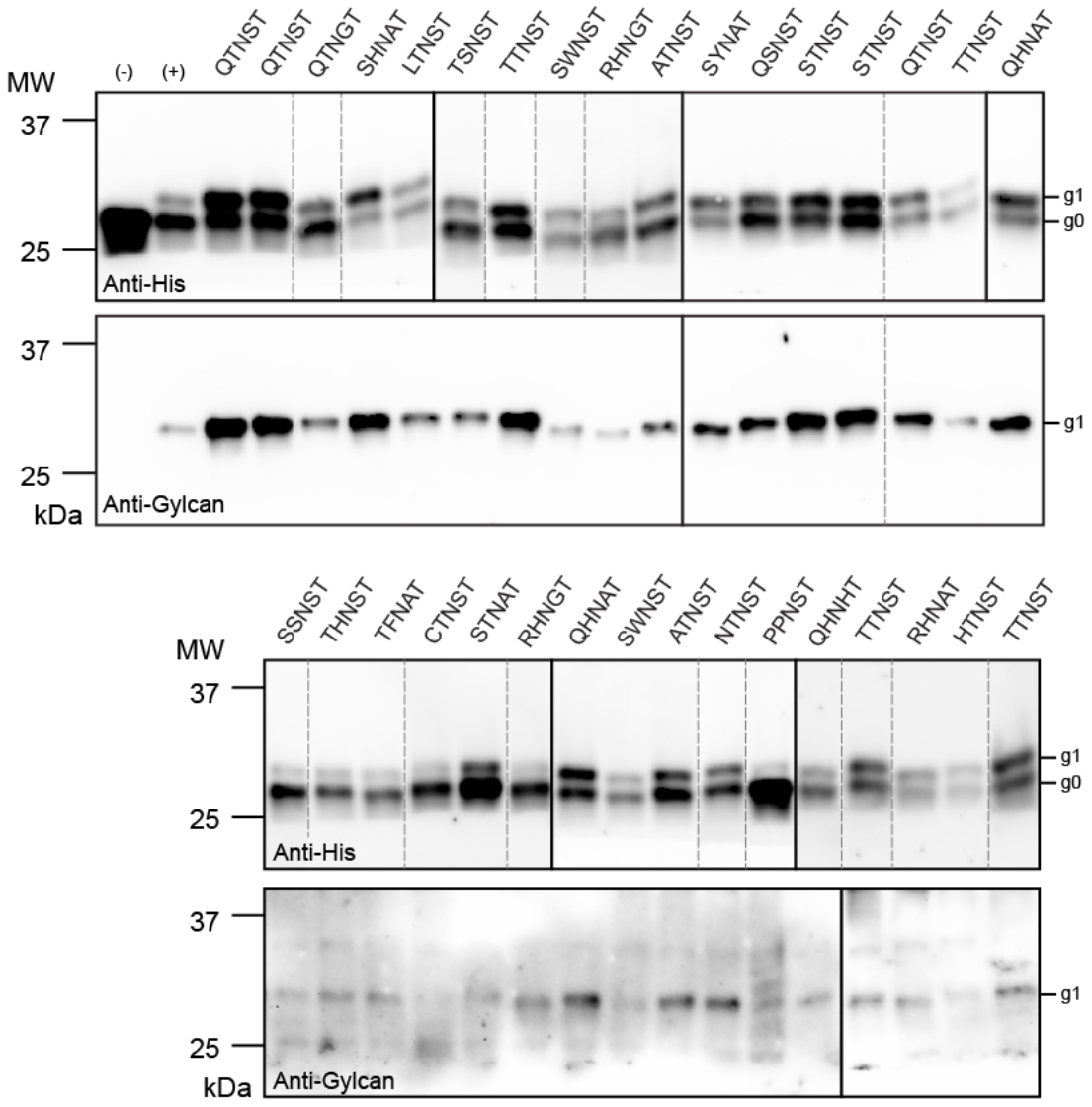

Figure 7. Positive hits of DdPglB against YebF-Im7^{XXNXT} library. Immunoblot analysis of periplasmic fractions from CLM24 cells complemented with DdPglB, and co-expressing the biosynthetic pathway for the *C. jejuni* glycan and protein YebF-Im7^{XXNXT} (X indicate random amino acids). Samples were picked from positive hits of the glycoSNAP assay. Sequencing results are shown on top of the blot. Molecular weight (MW) markers are indicated on the left. The g0 and g1 on the right indicate aglycosylated and glycosylated forms of the target protein, respectively. Black boxes denote the samples in the same blot, while gray dotted lines denote the samples are nonadjacent.

From the results in Figure 7, we observed most of the samples' glycosylation efficiencies are around 50% or lower. The positive control DdPglB against sequon AQNAT had an efficiency of 30%. One of the noticeable sequons was SHNAT, which had the highest glycosylation efficiency of 70% among this panel.

After the thorough screening and Western blotting verification, we collected several positive hits of sequon which can be glycosylated by DdPglB. Next, the sequence logo of DdPglB against XXNXT sequon was generated by using Weblogo 3[34]. The result is shown in Figure 8. The result indicated DdPglB did not have strong preferences at -2 position of the acceptor motif compared to CjPglB, which required D/E at -2 position. A preference of S/A/G at +1 position was also observed in this result. Overall, from this sequence logo, DdPglB did have a relaxed acceptor-site specificity.

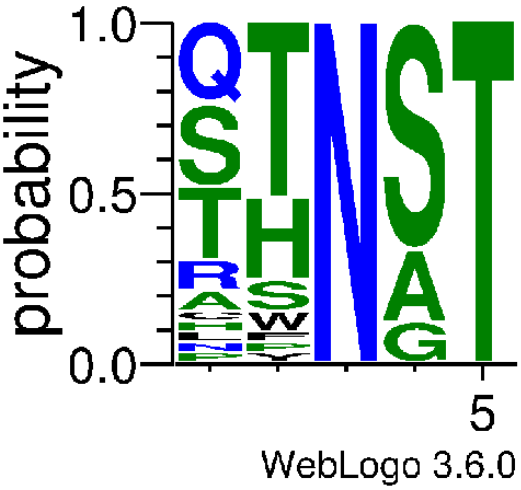

Figure 8. Sequence logo of DdPglB against YebF-Im7^{XXNXT} library. Sequence logo showing experimentally determined acceptor-sequon specificity of DdPglB against YebF-Im7^{XXNXT} library using glycoSNAP screening assay.

Chimeric *C. jejuni-D.* desulfuricans OST

After the investigation of relaxed sequon specificity of DdPglB, engineered OSTs were then constructed and tested for substrate specificity. Engineered OSTs were constructed by standard cloning technic which combined transmembrane domain and periplasmic domain of two different OSTs, CjPglB and DdPglB. DdPglB has been verified for its relaxed specificity while CjPglB was chosen for its high glycosylation efficiency. Engineered chimeric OSTs that combined transmembrane domain of CjPglB (amino acids numbered 1-429 of CjPglB) with periplasmic domain of DdPglB (amino acids numbered 469-729 if DdPglB), $Cj^{TM}Dd^{PD}$OST, and a reverse mutant $Dd^{TM}Cj^{PD}$OST (combined amino acids numbered 1-468 of DdPglB and 430-714 of CjPglB) were then tested against the acceptor motif XQNAT where -2 position was varied to include all 20 amino acids. The same *C. jejuni* glycan production pathway and acceptor protein YebF-Im7 were used in this experiment. Results are presented below in Figure 9.

From the results, we noticed that while the proteins were expressed well in the immunoblot analysis, both engineered $Cj^{TM}Dd^{PD}$OST and $Dd^{TM}Cj^{PD}$OST did not have the ability to glycosylate any of the sequons with 20 amino acids at -2 position. Compared to the previous study in Figure 4, we noticed their original ability of glycosylating certain sequon was inactivated after the combination.

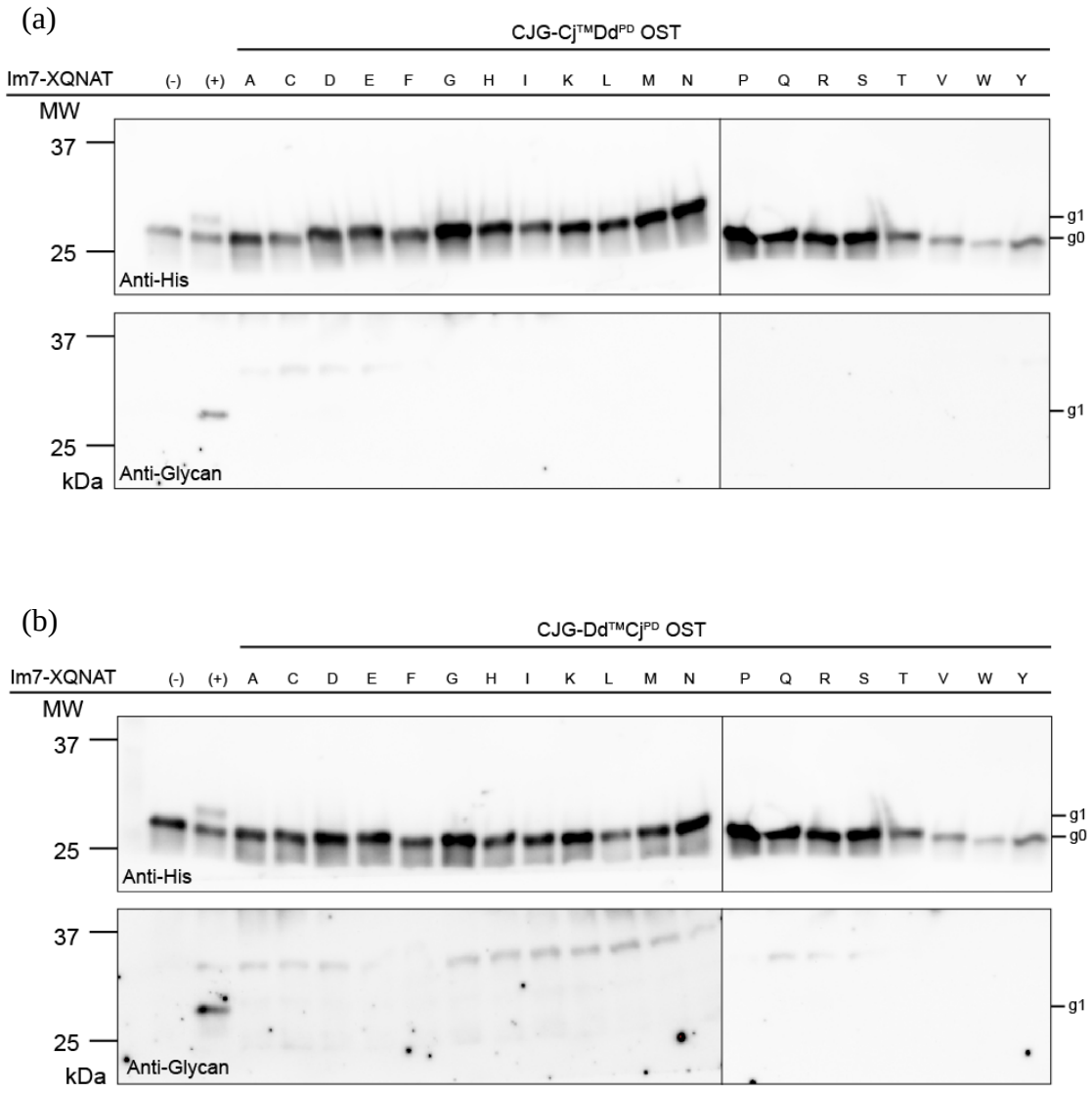

Figure 9. Specificities of chimeric OSTs combined CjPglB and DdPglB. Immunoblot of acceptor protein YebF-Im7^{XQNAT}, where X is one of the 20 amino acids indicated on top, co-expressed with the biosynthetic pathway for the *C. jejuni* glycan (CJG) and engineered (a) Cj^{TM}Dd^{PD}OST and (b) Dd^{TM}Cj^{PD}OST. Molecular weight (MW) markers are indicated on the left. The g0 and g1 on the right indicate aglycosylated and glycosylated forms of the target protein, respectively. Black boxes denote the samples in the same blot.

Glycosylation of native human IgG1 acceptor sequon by DdPglB

From the sequon logo result of DdPglB, we noticed there was an interesting potential sequon that can be glycosylated by DdPglB, QYNST, which is a single conserved glycosylation site of the human immunoglobulin G (IgG) on its Fc domain. Therefore, we performed an experiment to verify if DdPglB can glycosylate this sequon in our system. The plasmid containing acceptor protein YebF-Im7^{XQNAT} was then constructed and co-expressed with DdPglB and *C. jejuni* glycan pathway. The result is shown in Figure 10, which indicates sequon QYNST can be successfully glycosylated by DdPglB with 40% efficiency.

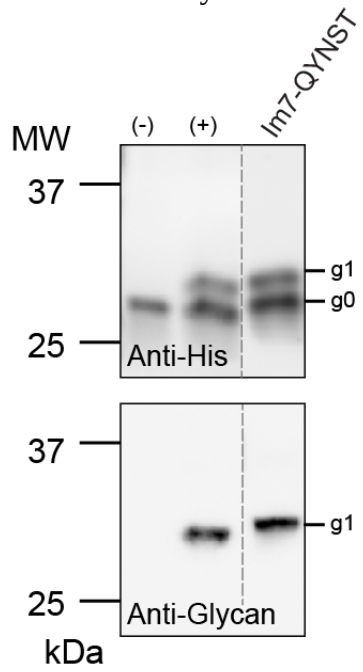

Figure 10. Glycosylation of DdPglB against YebF-Im7^{QYNST}. Immunoblot of protein YebF-Im7^{QYNST} co-expressed with the biosynthetic pathway for the *C. jejuni* glycan (CJG) and DdPglB. Molecular weight (MW) markers are indicated on the left. The g0 and g1 on the right indicate aglycosylated and glycosylated forms of the protein respectively. Dotted lines denote the samples are in the same blot but nonadjacent.

CHAPTER 4

DISCUSSION

DdPglB has been one of the OSTs with relaxed sequon specificity at the -2 position relative to the central asparagine. Here, we performed a more thorough investigation by the glycoSNAP screening against XXNXT sequon library, and the screening results indicated the acceptor-site specificity of DdPglB was certainly quite relaxed, especially at the -2 and -1 position, while DdPglB might have a preference for S/A/G amino acids at +1 position. The verification of the positive hits also provided a series of sequon that can be glycosylated by DdPglB and testified their performance on glycosylation efficiency. These results could serve as a glycoengineering toolbox and database for glycosylating various native protein acceptor sequons in the future. One noticeable result was the sequon SHNAT can be glycosylated by DdPglB with 67% glycosylation efficiency, which was the highest efficiency among all the literature reported efficiencies. Thus, here we also provided the sequon SHNAT as a new optimal acceptor sequon that was most suitable for glycosylation of DdPglB so far.

The single crossover engineered chimeric OSTs $Cj^{TM}Dd^{PD}$OST and $Dd^{TM}Cj^{PD}$OST were also constructed and investigated in this study. Transmembrane domain and periplasmic domains might have distinct abilities to support the reaction of OSTs, such as stabilizing the acceptor sequon or glycan complex by van der Waals contact, salt bridge, etc[25]. However, our results interestingly revealed that these two constructs could lose their ability to glycosylate the acceptor sequon that their original

full-length PglB were able to glycosylate. These indicated that some of the amino acids around the crossover point that we replaced in this experiment could play an important role in affecting the activity of these enzymes and their glycosylation ability. In the future, a series of single crossover chimeric OSTs could be constructed and screened through glycoSNAP to investigate what section of PglBs is the key part that affects glycosylation activity. This study and the future plan might provide an insight into the glycosylation mechanism of OSTs.

In this study, we also confirmed that sequon QYNST with the model protein Im7 can be glycosylated by DdPglB in *E coli*. QYNST is the native conserved glycosylation site at N297 of the Fc domain of human IgG. Therefore, successful glycosylation of this sequon has great significance here. So far, there are no other bacteria OSTs proved to have the ability to glycosylate sequon QYNST with efficiency higher than that in this study, which is 40%. In the future, we can perform a further investigation to check if DdPglB can also glycosylate the native Fc domain of human IgG, which might have the potential to be expressed as the glycoprotein with high utilization.

CHAPTER 5

CONCLUSION

This study provided a thorough investigation into the acceptor-site preference of DdPglB, and demonstrated DdPglB had a quite relaxed sequon specificity by building the sequence logo. A series of sequons which can be glycosylated by DdPglB was also presented, as well as their glycosylation performance. Thereby, the database of potential acceptor-sequon was expanded.

In addition, the engineered chimeric OSTs $Cj^{TM}Dd^{PD}OST$ and $Dd^{TM}Cj^{PD}OST$ were constructed here. Their glycosylation activity was eliminated after the combination of transmembrane and periplasmic domains from the two different bacterial PglBs. This indicated amino acids around the crossover section might serve a key role in the functionality of these OSTs.

Finally, sequence logo of DdPglB indicated QYNST, a native acceptor sequon of Fc domain of human IgG, could be potentially glycosylated by DdPglB. The result was also verified by immunoblot analysis and proved that DdPglB was able to glycosylate this sequon attached on C-terminus of the model protein YebF-$Im7^{QYNST}$ with an efficiency of 40%. Our research showed that DdPglB had the potential for glycosylating the native acceptor-site of human IgG.

In conclusion, we presented a list of sequons that can be glycosylated by DdPglB and further expanded the glycoengineering toolbox with more potential acceptor sites.

APPENDIX

Supplementary information

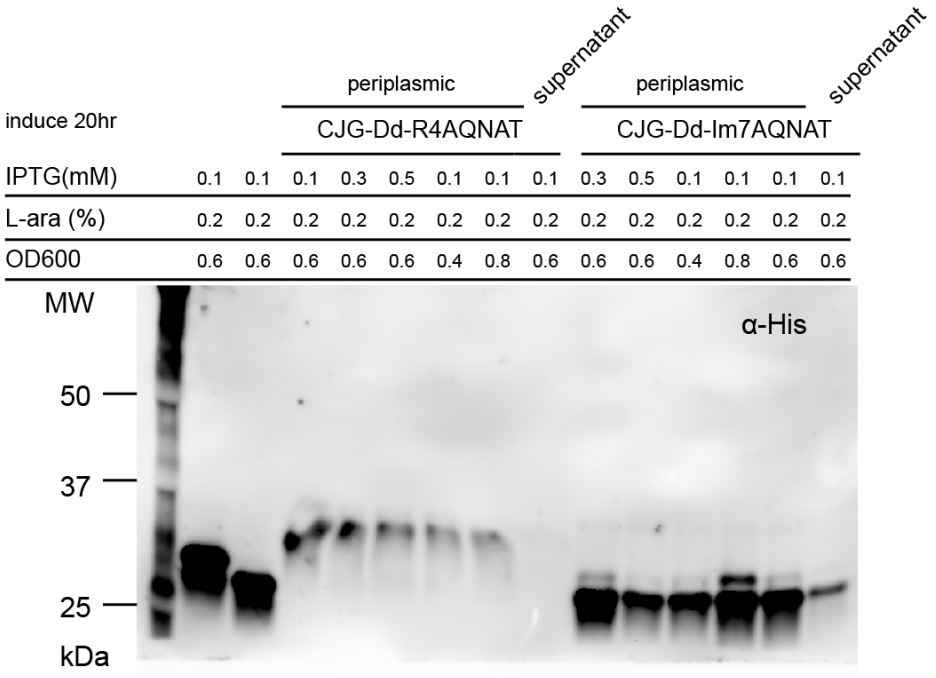

Figure S1. Expression optimization of R4^{AQNAT} and Im7^{AQNAT} at different conditions.

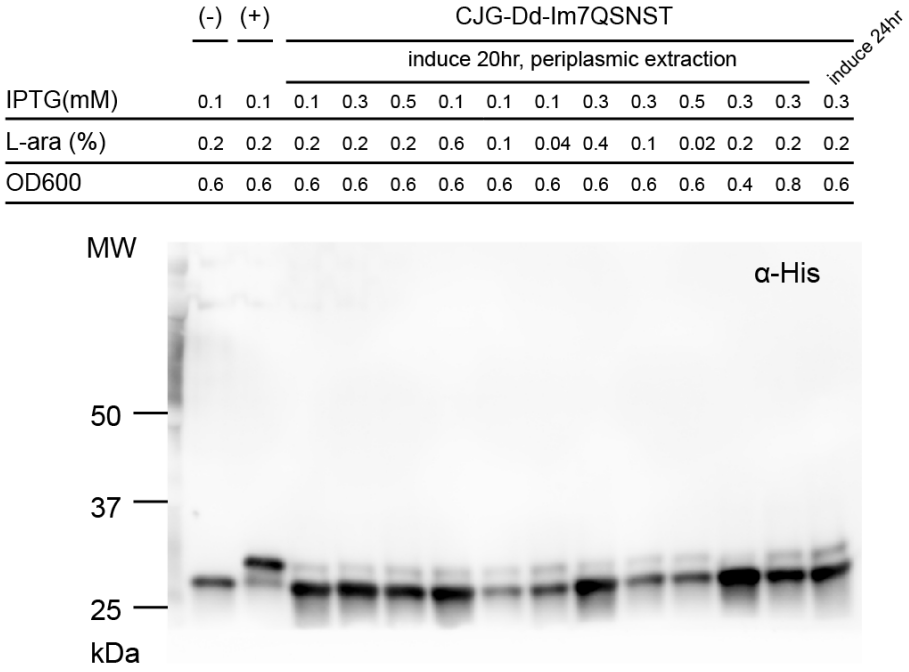

Figure S2. Expression optimization of Im7^{QSNST} at different conditions.

	(-)	(+)	CJG-Dd-Im7AQNAT											
			induce 20hr, periplasmic extraction											
IPTG(mM)	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.1	0.1	0.3	0.1	0.1	0.1	0.3
L-ara (%)	0.2	0.2	0.04	0.1	0.2	0.2	0.04	0.1	0.2	0.2	0.04	0.1	0.2	0.2
OD600	0.6	0.6	0.4	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.8	0.8	0.8	0.8

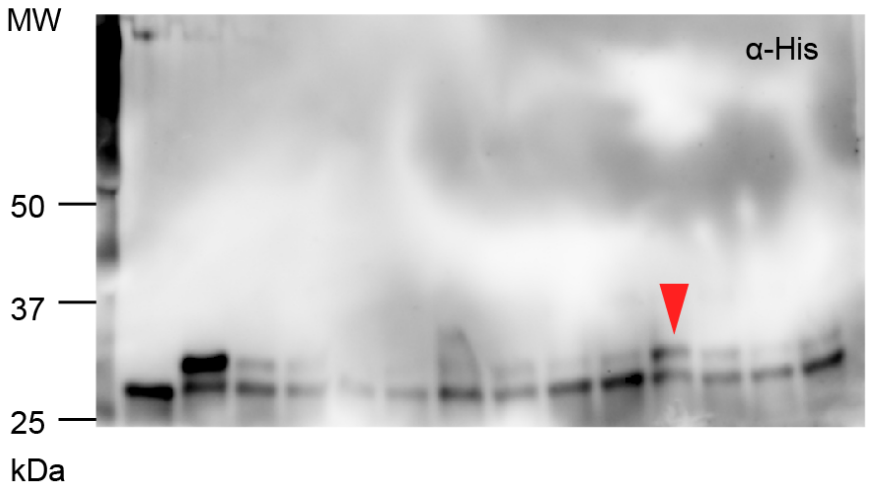

Figure S3. Expression optimization of Im7^{AQNAT} at different conditions. (Red arrow indicates the optimize condition)

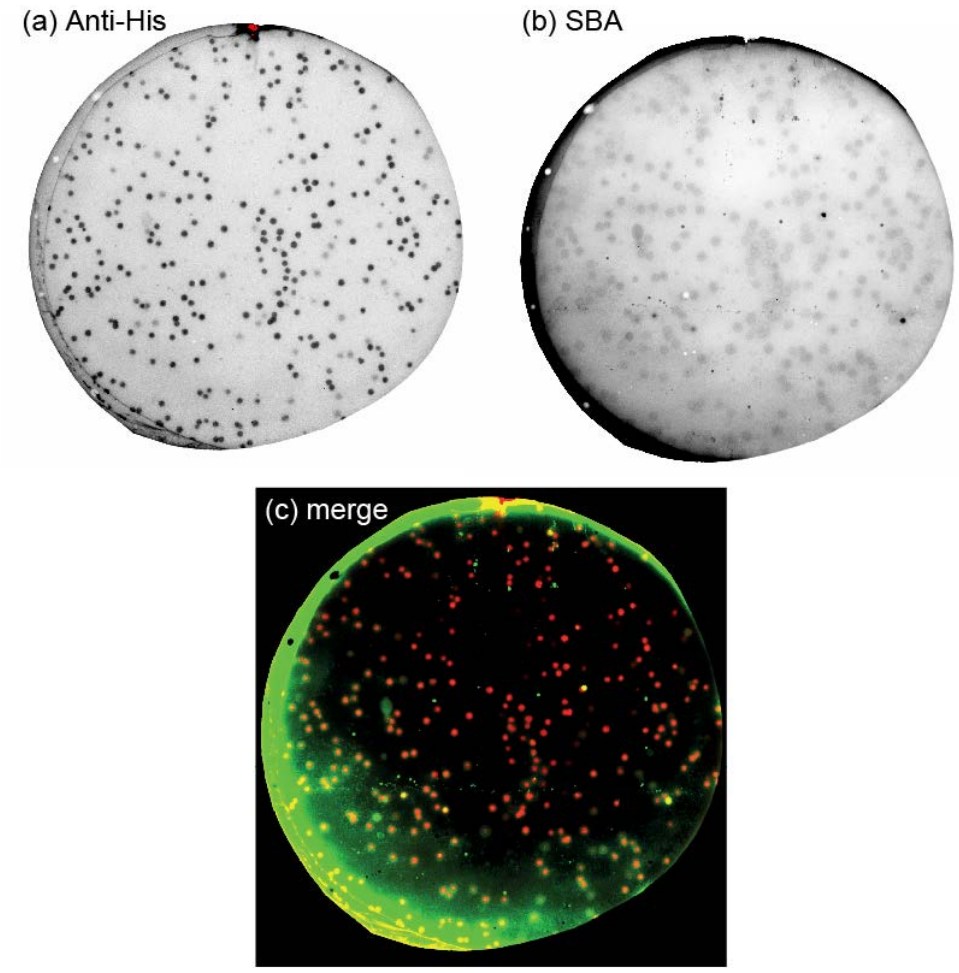

Figure S4. Exemple of glycoSNAP result imaging by ChemiDoc.

REFERENCES

1. Apweiler, R., Hermjakob, H. &Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
2. Castilho, A. Glyco-engineering: Methods and protocols. *Glyco-Engineering Methods Protoc.* **1321**, 1–439 (2015).
3. Helenius, A. &Aebi, M. Intracellular functions of N-linked glycans. *Science* **291**, 2364–9 (2001).
4. Imperiali, B. &O'Connor, S. E. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* **3**, 643–649 (1999).
5. Helenius, A. &Aebi, M. Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
6. Aebi, M., Bernasconi, R., Clerc, S. &Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends Biochem. Sci.* **35**, 74–82 (2010).
7. Defaus, S., Gupta, P., Andreu, D. &Gutiérrez-Gallego, R. Mammalian protein glycosylation – structure versus function. *Analyst* **139**, 2944–2967 (2014).
8. Wolfert, M. A. &Boons, G.-J. Adaptive immune activation: glycosylation does matter. *Nat. Chem. Biol.* **9**, 776–84 (2013).
9. Agard, N. J. &Bertozzi, C. R. Chemical approaches to perturb, profile, and perceive glycans. *Acc. Chem. Res.* **42**, 788–97 (2009).
10. Sinclair, A. M. &Elliott, S. Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *J. Pharm. Sci.* **94**, 1626–1635 (2005).
11. Solá, R. J. &Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *J. Pharm. Sci.* **98**, 1223–1245 (2009).
12. vanKooyk, Y. &Rabinovich, G. A. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat. Immunol.* **9**, 593–601 (2008).
13. Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).
14. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. &Guerry, P. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol. Microbiol.* **32**, 1022–30 (1999).
15. Nothaft, H. &Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **8**, 765–778 (2010).
16. Jervis, A. J. *et al.* Characterization of the Structurally Diverse N-Linked Glycans of Campylobacter Species. *J. Bacteriol.* **194**, 2355–2362 (2012).

17. Weerapana, E. &Imperiali, B. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R-101R (2006).

18. Baker, J. L., Çelik, E. &DeLisa, M. P. Expanding the glycoengineering toolbox: the rise of bacterial N-linked protein glycosylation. *Trends Biotechnol.* **31**, 313–323 (2013).

19. Yan, Q. &Lennarz, W. J. Studies on the Function of Oligosaccharyl Transferase Subunits. *J. Biol. Chem.* **277**, 47692–47700 (2002).

20. Schwarz, F. &Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21**, 576–582 (2011).

21. Gavel, Y. &Heijne, G.von. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *"Protein Eng. Des. Sel.* **3**, 433–442 (1990).

22. Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S. &Aebi, M. The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **15**, 361–367 (2005).

23. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966 (2006).

24. Chen, M. M., Glover, K. J. &Imperiali, B. From Peptide to Protein: Comparative Analysis of the Substrate Specificity of N-Linked Glycosylation in *C. jejuni* †. *Biochemistry* **46**, 5579–5585 (2007).

25. Lizak, C., Gerber, S., Numao, S., Aebi, M. &Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350–355 (2011).

26. Napiórkowska, M., Boilevin, J., Darbre, T., Reymond, J.-L. &Locher, K. P. Structure of bacterial oligosaccharyltransferase PglB bound to a reactive LLO and an inhibitory peptide. *Sci. Rep.* **8**, 16297 (2018).

27. Ollis, A. A., Zhang, S., Fisher, A. C. &DeLisa, M. P. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat. Chem. Biol.* **10**, 816–822 (2014).

28. Ihssen, J. *et al.* Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol.* **5**, 140227 (2015).

29. Ollis, A. A. *et al.* Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci. Rep.* **5**, 15237 (2015).

30. Wacker, M. *et al.* N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into E. coli. *Science (80-.).* **298**, 1790–1793 (2002).

31. Ollis, A. A., Chai, Y. &DeLisa, M. P. GlycoSNAP: A High-Throughput Screening Methodology for Engineering Designer Glycosylation Enzymes. in 37–47 (Humana Press, New York, NY, 2015). doi:10.1007/978-1-4939-2760-9_3

32. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc. Natl. Acad. Sci.* **102**, 3016–3021 (2005).

33. Krogh, A., Larsson, B., vonHeijne, G. &Sonnhammer, E. L. . Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *J. Mol. Biol.* **305**, 567–580 (2001).

34. Crooks, G. E., Hon, G., Chandonia, J.-M. &Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–90 (2004).